HOSTSEQ

DATABASE GOVERNANCE FRAMEWORK

August 4, 2020.

TABLE OF CONTENTS

Acronyms **2**
Definitions **3**
1. Overview **4**
1.1. Objective and general description of the HostSeq Databank 4
1.2. Purpose of this Governance Framework 4
2. Governance and organizational structure **5**
2.1. Governance overview 5
2.2. Scientific and administrative oversight 6
2.3. Ethical oversight 6
2.4. Data access and release oversight 6
2.5. Management of the database 7
2.6. Financing 7
3. Types of data included in the HostSeq Databank **7**
3.1. Overview 7
4. Data collection and deposit in HostSeq **9**
4.1. Process for depositing data from local research collections 9
5. Data management **11**
5.1. Overview of data flow in HostSeq 11
5.2. Storage of data 12
5.3. Data access 12
5.3.1. Open access data 14
5.3.2. Controlled-access data 14
5.4. Data sharing with other databases 14
6. Database closure **14**
7. Potential benefits and risks to research participants **15**
7.1. Benefits 15
7.2. Physical risks 15
7.3. Informational risks 15
8. Protection of participant privacy and data security measures **15**
9. Duration of conservation of data **16**
10. Management of participant withdrawal **16**
11. Participant re-contact **17**
12. Return of research results and incidental findings **17**
13. Results and intellectual property **17**
Appendix A: Core Research Consent Elements for prospective collection of datasets **19**
Appendix B: Example of consent clauses for the prospective Core Consent Elements **21**
Appendix C: Assessment of consent elements for inclusion of datasets from retrospective collections **23**
Appendix D: Data Access Compliance Office (DACO) Terms of Reference **25**
Appendix E: Steps to request access to controlled-access datasets **27**

Acronyms

COVID-19: Coronavirus disease 2019
DACO: Data Access Compliance Office
REB: Research Ethics Board
TCAG: The Centre for Applied Genomics, the Hospital for Sick Children Toronto
TCPS: Tri-Council Policy Statement
WGS: Whole genome sequencing

Definitions

Access: To retrieve, copy or process a digital, conceptual or physical asset (including a dataset), in whole or in part.

Approved Third-Party Researcher: A researcher who is not affiliated with HostSeq, CGEn or a Contributing Collection (i.e. is not listed as part of study team members) and has been approved by the HostSeq Data Access Committee to access HostSeq datasets.

Biological samples: DNA or blood samples

Collection-affiliated Researcher: A researcher who is directly affiliated with a collection from which data has been contributed/deposited to HostSeq ("Contributing Collection"). Such researcher is part of the collection's research team (principal investigator, co-investigator or authorized research team member).

Contributing Collection: Means a group of one or more individuals ("Participants") whose Biological Samples and data and/or samples have been collected as part of a research project at a participating Canadian institution or other Canadian research infrastructure and will be transmitted to a CGEn site for sequencing and deposit in the HostSeq databank.

Prospective Collection: A potential Contributing Collection that is being collected and consented specifically for inclusion in the HostSeq databank.

Retrospective Collection: A potential Contributing Collection that has been collected or is being collected under REB permission obtained prior to the creation of the HostSeq databank, or that for other reasons, the REB permission and/or consent does not include consideration of HostSeq.

1. Overview

1.1. Objective and general description of the HostSeq Databank

In light of the ongoing COVID-19 pandemic, research on SARS-Cov-2 (COVID-19) will be essential to better understand the evolution of this novel strain of coronavirus. It will serve to identify genetic risk factors for infection and take appropriate protective measures and aid health authorities in adopting social measures to control the spread of infection and, ultimately, serve to better anticipate and prepare for future pandemic outbreaks.

To foster research on COVID-19, CGEn and Genome Canada have mandated a group of researchers to create the HostSeq Databank ('**HostSeq**').

In this context, the goal of the HostSeq Databank ('**HostSeq**') is to create a unique, collaborative research infrastructure whose mission is to conduct the whole genome sequencing ("**WGS**") of 10,000 human genomes using Biological Samples provided by research participants susceptible to have been infected with the novel coronavirus SARS-CoV2 and enable data sharing on COVID-19 and other health outcomes. HostSeq will focus on collecting WGS data of affected individuals (hosts), however, when possible, HostSeq will link COVID-19 viral sequence information to host datasets (this could be done, for example, through access to health/administrative records, through consent to sharing of data between studies, etc.). This will help the scientific and medical communities to better understand, treat and ultimately limit the impact of COVID-19. This infrastructure will be made available to researchers in Canada and internationally, and will aim to collaborate and be interoperable with other similar resources to maximize opportunities to combine shareable datasets.

HostSeq will include data from numerous Contributing Collections, both prospective and retrospective. While each Contributing Collection has a fiduciary duty over the samples and data it contributes to HostSeq, CGEn will assume a custodianship role over the centralized databank through its host instruction, the Hospital for Sick Children ("**SickKids**"). HostSeq will make data available for research on COVID-19 and other health outcomes.

Data housed in HostSeq will be stored on servers located in Canada, under the custodianship of SickKids, its host institution. Local research projects and investigators will be able to contribute Biological Samples and data to HostSeq if they meet conditions described in this governance framework.

Finally, HostSeq may also collaborate with other international COVID-19 initiatives, by identifying appropriate data sharing opportunities, in order to maximize international collaboration and knowledge on the virus.

1.2. Purpose of this Governance Framework

This Governance Framework aims to provide oversight mechanisms regarding the administration, custodianship and sharing of data deposited in HostSeq. In addition, it sets out the criteria for the inclusion of data from eligible research collections (prospective and retrospective).

Finally, this Framework sets out ethical considerations pertaining to the deposit and storage of data in HostSeq (e.g. consent requirements, description of privacy and security protections, closure/transfer of the database, management of participant withdrawal, risks and benefits, etc.).

This Governance Framework is intended to assist potentially eligible collections and their local **REB** in assessing the oversight mechanisms in place to ensure appropriate data sharing via the HostSeq Databank. It also informs researchers accessing the resource on the different types of data available, and the overall access process.

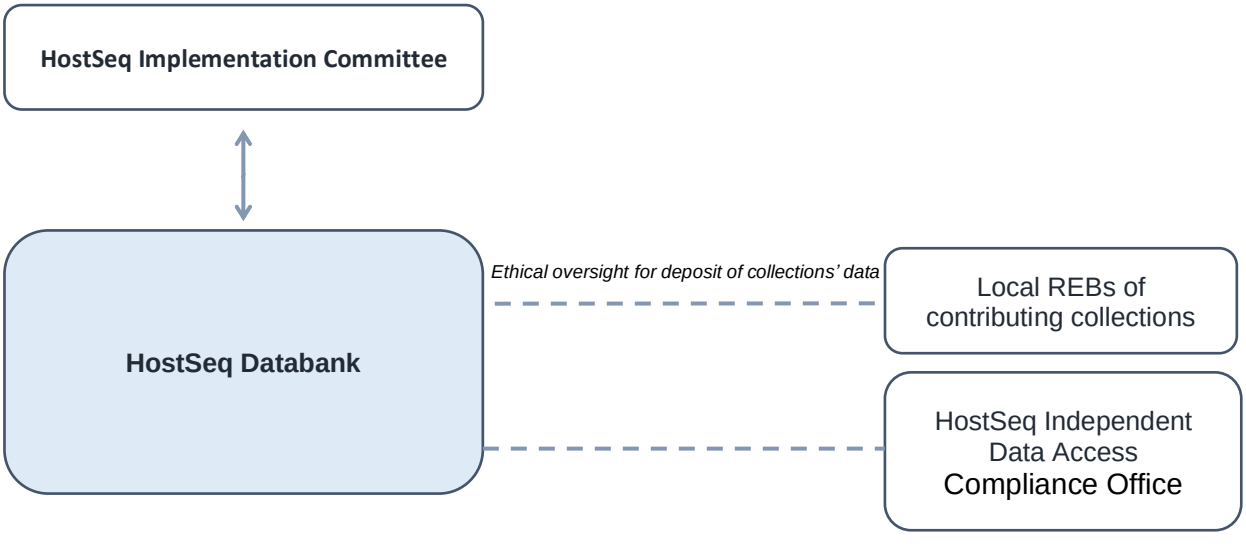
1.3. Mandatory Commitments

As part of contributing samples and data to HostSeq, researchers should note that the following commitments are mandatory (and will be part of institutional agreements):

Minimal contribution requirements from contributing researchers/institutions:
• A copy of sequence data will be added to the HostSeq Databank (contributing research team will receive their own copy of this data) • Relevant clinical/phenotype data will need to be provided and will be stored in the HostSeq Databank • Access to the full HostSeq Databank (including genomic and clinical data contributed) by approved researchers will be governed through applications to the HostSeq Data Access Compliance Office (DACO)

2. Governance and organizational structure

2.1. Governance overview



2.2. Scientific and administrative oversight

The HostSeq Implementation Committee provides scientific and strategic orientations for the HostSeq Databank, in order to optimize scientific outcomes and benefits to Canadians.

2.3. Ethical oversight

The HostSeq Database will include data from collections contributed by Collection-affiliated Researchers.

For Retrospective Collections, additional approval will be sought from the local institutional REB to allow Biological Samples to be analyzed and data from the collection's participants to be included in HostSeq, according to the conditions established in this Governance Framework (see Section 4.1).

For Prospective Collections, local REB approval for analysis of Biological Samples and inclusion of participants' data in the HostSeq should be sought at the time of initial ethics review submission to the local REB responsible for ethics oversight of that collection (see Section **Error! Reference source not found.**), and information on the contribution to the HostSeq Databank should be provided in the participant consent forms.

Finally, ethics approval will be sought from the SickKids REB, as the board of record for the overall creation of the HostSeq infrastructure.

2.4. Data access and release oversight

Data from the HostSeq Databank as described in Table 1 will be made available through open, and controlled-access tiers, based on the type of data, its sensitivity and the specification of the contributing collection.

Data access and release under the controlled access tiers will be overseen by the HostSeq Independent Data Access Compliance Office ('**DACO**'), as described in Section 5.3.

The DACO has adopted its own terms of reference as detailed in Appendix D (including, for instance, the frequency of its meetings, the process for the review of access applications, etc.). It shall be composed of a minimum of five (5) voting members, independent from CGen (the sponsor), including:

- At least one (1) infectious disease expert;
- At least one (1) bioinformatics researcher [voting member];
- At least one (1) expert in the technical infrastructure of the HostSeq Databank [voting member];
- At least one (1) expert on the legal/ethical aspect of genomic research, data sharing, privacy and data protection [voting member];
- At least one (1) patient/participant representative or member of the public [voting member];
- One (1) investigator representing the HostSeq Databank [non-voting member].

For certain access applications, the DACO may call upon outside experts. Such expert will not be a voting member on the Committee, but invited to provide background expertise required to review the application.

2.5. Management of the database

The HostSeq Databank will be managed by a Program Manager, who will oversee research administration, including finance, contracts, operations, staffing, and communication.

2.6. Financing

The HostSeq Databank is financed by CanCOGeN funding provided to CGEn through Genome Canada.

Ongoing funding sources will be sought to ensure the financial sustainability of the HostSeq Databank.

In the future, a cost-recovery structure may be adopted, to ensure sustainability of the resource (for example, access fees may be charged to certain users or for certain uses of computing resources). This fee structure will be submitted to the SickKids REB prior to implementation.

3. Types of data included in the HostSeq Databank

3.1. Overview

HostSeq will receive Biological Samples and data from existing and future COVID-19 research studies acting as Participating Collections. The following sections provide an overview of the types of data that can currently be included in HostSeq (this list may be updated, in time).

Participating collections will be required to provide a defined set of phenotypic data using World Health Organization ISARIC approved data forms, with some information specific to COVID-19 co-morbidities, as well as a blood or DNA sample for whole genome sequencing at the Centre for Applied Genomics (TCAG), at SickKids, or another CGEn site. The phenotype data will be entered into a REDCap Database by the submitting investigator (or designate) from local contributing collections or sent via email and password-protected spreadsheets to be entered at the Centre for Applied Genomics (TCAG).

Other types of data are optionally contributed by Participating Collections in agreement with the HostSeq Scientific Implementation Committee, based on availability, quality, as well as any applicable consent or data use requirements. These data will be associated with and linked to an individual-level participant HostSeq record.

Table 1: List and description of data available in the HostSeq Databank, and minimal collection data deposit requirements as listed in the HostSeq case report form/intake form (available online at: https://redcapexternal.research.sickkids.ca/surveys/?s=8CXDYF9AJ3).

The datasets included in HostSeq will depend on a local data collection practices and consent permissions (where applicable). Not all data types listed in this table will be collected by every collection or for every participant.

(i) **PHENOTYPIC DATA**	**Description**	
a. Phenotype data from research collections	Clinical phenotype information will be requested from the collections, largely based on World Health Organization ISARIC approved data forms, with some information specific to COVID-19 co-morbidities, and information to enable data linkages.	Required
Linkage with administrative datasets	Where implemented by the local collection, linkage to certain administrative datasets (e.g. electronic healthcare records, admissions and cost data, radiological and imaging examinations, physical assessment, general health.)	Required
(ii) **HOST TISSUE SAMPLE/ MOLECULAR DATA**	**Description**	
a. Biological sample	A Biological Sample will be provided by the local collection, for whole genome sequencing (WGS) by CGEn. Biological Samples are sent to one of three CGEn nodes for DNA extraction. The DNA is sequenced at the designated CGEn node. These samples will be linked to WGS data through the RedCap database.	Required
b. WGS data	HostSeq will include datasets derived from whole genome sequencing. If available, these datasets can be contributed by the local collection. All genomic datasets will be analyzed using the standardized HostSeq analysis pipeline (as detailed in the HostSeq Databank Protocol).	
(iii) **COVID-19 VIRAL SEQUENCING DATA**	**Description**	
a. Viral sequencing data	If available, individual-linked COVID-19 viral sequencing data can be included in the HostSeq Databank.	Optional

4. Data collection and deposit in HostSeq

The HostSeq Databank will accept datasets from Prospective Collections specifically consented for HostSeq as well as samples for sequencing and datasets from Retrospective Collections not specifically consented for HostSeq, provided they meet the core consent elements identified below.

Recruitment Sites

Participants will be recruited at different participating sites, across Canada. The local site must have ethics approval in place for the project collecting samples or data in relation to COVID-19.

Identification of Potential Participants

The Principal Investigator (or delegate) of the Contributing Collection ensures that potential participants are identified for data inclusion in HostSeq, in accordance with the present Governance Framework, as well as all applicable laws, guidelines, norms and institutional policies.

Where participants are identified through their medical records, certain additional institutional approvals may be required to access medical records for screening purposes. In certain institutions, the treating physician may participate in the identification of potential participants, where approved as a recruitment method.

4.1. Process for depositing data from local research collections

*4.1.1. **Prospective collections***

Consent to the collection of Biological Samples by local sites, followed by the genetic analysis of Biological Samples and sharing of research data through the HostSeq Databank, should be obtained by researchers at the local site. Individual who meet the eligibility criteria listed above can be approached for consent and enrolled in the study.

Consent for linkage with administrative datasets (including the participant's clinical data) will be obtained in order for research staff to:

1. Collect and use relevant information from the participant's medical record that is related to HostSeq's objectives; and
2. Access the participant's data from other administrative or governmental databases;
3. Where possible, access to the participant's data from other research projects (e.g. viral genome data).

Due to the nature of COVID-19, in many cases, consent may need to be obtained verbally, and necessary adjustment to the consent form and consent guidelines should be implemented, as recommended by the local site. For example, should consent be obtained verbally, local sites could choose to contact potential participants by telephone or by other means of communication, present the HostSeq Databank and answer any potential queries, in accordance with local procedures. In the event that a potential participant lacks the capacity to consent or has been rendered incapable due to a sudden and temporary incapacity, research personnel should contact the individual's legal

representative for verbal consent. If a person who has been rendered incompetent due to a sudden and temporary incapacity regains capacity, the standard procedures for obtaining consent from a capable person will be followed, according to local site policies.

Prospective collection consent forms should minimally include the following **core consent elements** to foster broad data sharing through the HostSeq Databank[1]:

	If a researcher at a local site wishes to deposit datasets in **HostSeq COVID-19 controlled-access Databank,** research consent should be obtained for:
Research data	Whole genome sequencing of the sample, existing clinical data and the ongoing collection of clinical data from participant's medical records/chart, administrative databases, etc.
International sharing	International sharing of genetic and clinical data
Future research use	Future health research on COVID-19 and other health outcomes
Commercial use	Use of genetic and clinical data for commercial purposes
Controlled access	Sharing of genetic and clinical data through a controlled-access mechanism
Storage on cloud servers in Canada	Possible storage of genetic and clinical data in the HostSeq Databank, on centralized Canadian cloud servers
Duration of storage	Indefinite storage of genetic and clinical data
Data withdrawal	Not possible to withdraw data that has already been distributed and used
Re-identification	Low risk that the participant could be re-identified in the future
Recontact (not mandatory, strongly recommended)[2]	Optional recontact of participants (yes/no)

If any of the items listed above are not included in the consent documents, datasets should not be deposited as-is in the HostSeq Databank without obtaining appropriate approvals from the collection's local research ethics committee.

*4.1.2. **Retrospective datasets***

Retrospective datasets include for example:

- Datasets that could be generated from already collected tissue samples (e.g. archival samples, samples collected in the context of another research project/biobank, samples collected, etc.);
- Datasets that were generated before the creation of the HostSeq COVID-19 Databank.

[1] Examples of appropriate consent language for each of these core elements are provided in Appendix B.

[2] HostSeq strongly recommends that contributing collections include an option to re-contact participants in their consent forms. HostSeq will accept datasets from participants who have agreed or refused re-contact, however, the databank will track whether participants have agreed to be re-contact or not in order to determine whether future data/sample collection is possible.

Retrospective consent materials may have used different language and in some cases may be ambiguous, or be silent, as to sample and data sharing and their potential uses. To help researchers from local sites determine whether samples and/or data can be used for the HostSeq COVID-19 Database, an assessment tool was developed (Appendix C), based on the principles found in the Global Alliance for Genomics and Health's Consent Policy[3] and other best practice guidelines[4].

*4.1.3. **REB approval to deposit data to the HostSeq Databank***

Approval from the collection's local REB will be obtained prior to deposit of the collection's datasets in the HostSeq Databank, based on assessment of consent materials and data sharing permissions.

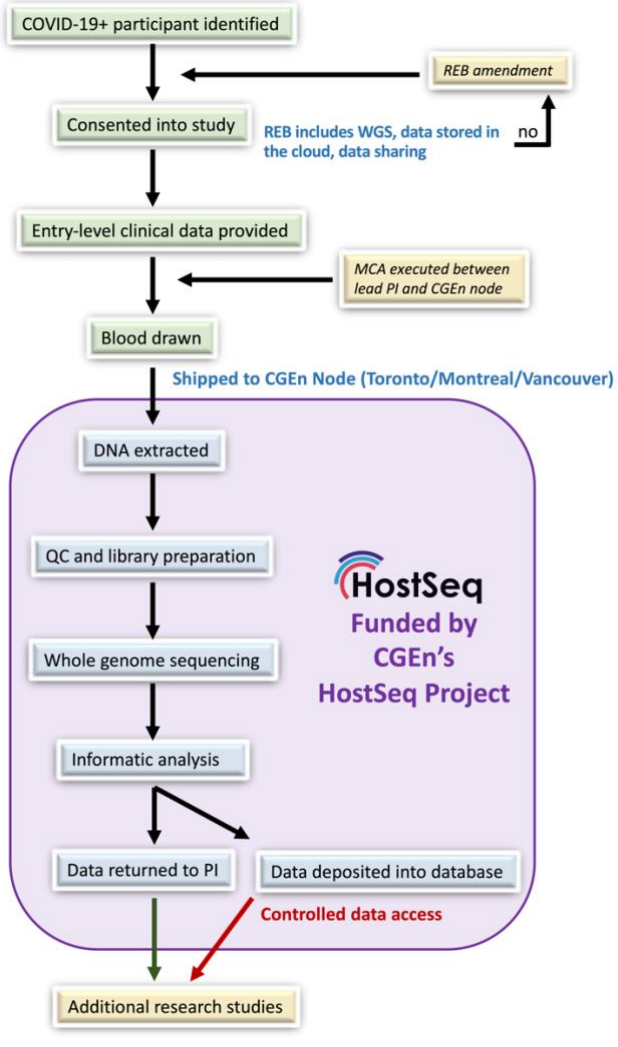
5. Data management

5.1. Overview of data flow in HostSeq

The following section summarizes the key components of the HostSeq data flow, from contribution of datasets by collections, to processing, to access and use. Each Participating Collection will have its own Standard Operating Procedure (SOP) for local data collection. However, a common data entry procedure has been developed for contributing to HostSeq, namely, data will be entered into a REDCap Database by the submitting investigator (or delegate) or sent via email and password-protected spreadsheets to be entered at the Centre for Applied Genomics (TCAG).

[3] Global Alliance for Genomics and Health: Consent Policy (September 2019), available online at: https://www.ga4gh.org/wp-content/uploads/GA4GH-Final-Revised-Consent-Policy_16Sept2019.pdf

[4] For example, Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans – TCPS2 (2018), Section 5.5A (Consent and secondary use of identifiable information for research purposes) and 12.3A (Consent and secondary use of human biological materials for research purposes).



5.2. Storage of data

Datasets in the HostSeq Databank will be hosted on local and/or cloud-based server located in Canada, under the custodianship of the host institution, SickKids.

Additional information will be provided as the Databank develops.

5.3. Data access

HostSeq data will be made available through open and controlled access tiers.

Table 2. Overview of HostSeq data access tiers

TIER 1: OPEN ACCESS	
Data available under Tier 1 (Open Access)	Data with a very low risk of re-identification and no particular sensitivity ("open data"), such as aggregated genomic and phenotypic data.
Who can access?	Openly accessible to anyone, through the use of APIs/HostSeq website (TBC)
Access mechanism	Open access
Access Process	TBD
TIER 2: CONTROLLED-ACCESS	
Data available under Tier 2 (Controlled Access)	Individual-level molecular and phenotypic data, including: • Phenotype data • Files in standard output file formats will be generated including gvcf/vcf files for SNV and indel variants, vcf files for SV and CNV calls and BAM and index files for visualization of the data. • Where available, associated COVID-19 viral sequencing data.
Who can access?	• Third-Party Researchers approved by the HostSeq DACO (as a clarification, Third-Party Research who have not submitted datasets can still apply to access the collection) • **Collection-affiliated Researchers (expedited review)**
Access mechanism	Controlled access
Access Process	**Third-Party Researchers**: Access to data by third-party researchers requires submission of an application to the HostSeq DACO (as detailed in Appendix E) Access requests must, amongst other, include: a summary of the research project; justification for the types of data requested; ethics approval from the researcher's institutional REB. **Collection-affiliated Researchers:** Access to data by Collection-affiliated Researchers requires submission of an expedited review application to the HostSeq DACO (as detailed in Appendix D) Access requests must, amongst other, include: a summary of the research project; justification for the types of data requested; ethics approval from the researcher's institutional REB.

5.3.1. Open access data

Aggregate data in the HostSeq Databank will be accessible to all researchers through dedicated use interfaces.

5.3.2. Controlled-access data

Access to participant-level detailed genomic and phenotypic data, such as clinical data, gvcf/vcf files for SNV and indel variants, vcf files for SV and CNV calls and BAM and index files; will be made available upon approval by the HostSeq DACO. In the case of a request by **a Third-Party Researcher** to access controlled-access data, a full research application needs to be submitted to the DACO, for review and approval (as detailed in Appendix E). In the case of **Collection-affiliated Researchers**, access requests will be reviewed following an expedited review process (as detailed in Appendix E). Access will require that a data access agreement be signed with the HostSeq host institution, SickKids.

The DACO terms of reference, including criteria for the review of controlled-access requests, are provided in Appendices D and E.

5.4. Data sharing with other databases

In order to enable international data sharing and create shared resources, HostSeq efforts will be made to share datasets with other international databases and/or biobanks.

In such cases, access to HostSeq resources will be will be contingent on signing of a data access agreement or collaboration agreement.

Sharing of datasets or samples from other international collections with HostSeq will also be permitted if it meets consent and regulatory requirements, and will also be contingent on signing of a data access agreement or collaboration agreement.

6. Database closure

Data housed in HostSeq is under the custodianship of the SickKids CGEn node, a Canada Foundation for Innovation (CFI) Major Science Initiative (MSI) national platform for genome sequencing and analysis.

If the HostSeq Databank were to cease activities at any time in the future, efforts will be made to transfer the coded data to a third party that agrees to comply with the HostSeq policies and the terms of participants' consent (as a note: transfer of server storage entities or location does not constitute a transfer of data custodianship). Prior to any transfer of data custodianship and responsibilities with respect to the maintenance of the HostSeq Databank, each contributing collections' Principal Investigators and local REB will be notified of such decision to transfer data and provided with the opportunity to accept the terms of transfer, or to withdraw data from collections under its purview. A decision made with regard to the transfer or closure of the database will involve the Genome Canada Board of Directors or other relevant body.

7. Potential benefits and risks to research participants

7.1. Benefits

There may be no direct benefits to the participants contributing data to the HostSeq Databank.

The HostSeq Databank will improve COVID-19-related data sharing world-wide, enabling the discovery of causes of diseases like COVID-19, by building the infrastructure and tools needed to improve research and, eventually, diagnosis and treatment.

7.2. Physical risks

The HostSeq Databank is not directly responsible for biospecimen sampling procedures or phenotype data collection. Sampling procedures and risks associated with such sampling are detailed as part of each contributing collection's protocol and consent forms.

There is no physical risk involved in taking part in the HostSeq Databank.

7.3. Informational risks

Much like fingerprints, it is possible to identify someone if certain data about that person are put together from different sources. While strict data security and privacy measures are in place by the HostSeq Databank to protect the participants' privacy, there is always a small risk that the participants' data may lead to their re-identification. Moreover, as technology advances, there may be new ways of re-identifying participants that cannot be foreseen today. Personal health information may affect the participant's insurability or employment, although some countries, like Canada, have enacted legislation to protect against genetic discrimination.

There is a possibility that researchers accessing the HostSeq Databank will be able to link HostSeq individual-level datasets to a participant's identity based on molecular and phenotypic data in the HostSeq Databank. However, upon accessing data from the HostSeq Databank, researchers will have to sign agreements that prevent them from intentionally attempting to re-identify participants.

Despite the mechanisms in place to keep all information secure and to protect participants' privacy, there remains a risk of a privacy breach. If a privacy breach occurs, the HostSeq Databank will contact the participants' collection to inform it of the breach and explain what data may have been compromised. It is the responsibility of the local collection to communicate this information on to participants.

8. Protection of participant privacy and data security measures

Each participating Collection has a fiduciary duty of maintaining the privacy and confidentiality of their participants' nominal information, i.e. information that can be directly or indirectly traced back to an individual participant. This information is stored on secured servers in each Participating Collection's internal database.

To protect participants' information, upon inclusion of collection data in the HostSeq Databank, data will be coded according to the local collection's procedure. If the local collection's REB's or institution requires data to be double-coded prior to depositing data in the HostSeq Databank double coding should be done by the collection, which will ultimately be responsible for providing this new, second code to the HostSeq Databank.

The only link between the participant's nominal information and the coded data stored in the HostSeq Databank is the participant code (participant ID). The coded data is thereafter imported and stored in the network part of the HostSeq Databank. The key to link the participant's HostSeq Databank number and collection participant ID (e.g. linking log) is held locally by the collection principal investigator. The HostSeq Databank will not collect or store information that directly identifies participants (e.g. names, contact information, healthcare numbers, etc.). Therefore, the possibility of obtaining nominal information on participants by consulting the shared part of the HostSeq Databank is very low.

Only the authorized members of the HostSeq team, Collection-Affiliated Researchers, and Third-Party Researchers approved by the HostSeq DACO will have access to authorized sets of coded phenotypic and genomic data. In the case of Collection-Affiliated Researchers, access to datasets other than their own collection will require DACO approval.

9. Duration of conservation of data

If consented to, data stored in the HostSeq Databank will be kept indefinitely or until no longer useful for research purposes. However, data can be removed from the HostSeq Databank, for example, if a Collection-affiliated Researcher informs the HostSeq Databank that a participant has withdrawn consent, that the collection has ended its activities (for example, yearly REB approval has not been obtained), that the collection's institution or REB has requested that data be removed, or in the event that the HostSeq Databank stops its activities (see Section 6).

10. Management of participant withdrawal

Contributing data to the HostSeq Databank is voluntary and participants will continue to receive the best available care, whether or not they decide to share data with the HostSeq Databank or not.

Participants have the right to withdraw data from the HostSeq Databank at any time and without providing any reason. Withdrawal can be implemented by informing the Collection-affiliated Researcher who is contributing to the HostSeq Databank. Upon withdrawal from the HostSeq Databank and/or the participating collection[5], the Collection-affiliated Researcher shall contact the HostSeq Databank to request withdrawal and that data be destroyed and removed from the HostSeq

[5] Some participants may not have directly consented to contributing their data to the HostSeq Databank (for example, if core consent elements are met, and if contribution is approved by the collection's REB). In this case, withdrawal from the contributing collection will generally also require the collection's Principal Investigator to request withdrawal of datasets from the HostSeq Databank.

Databank. However, data that has already been distributed for research analysis, cannot be removed or destroyed to preserve the scientific integrity of the analysis.

11. Participant re-contact

Re-contact of participants, if applicable, is managed by the investigators responsible for the individual Contributing Collections. **HostSeq strongly recommends that contributing collections include an option to re-contact participants in their consent forms. HostSeq will accept datasets from participants who have agreed or refused re-contact, however, the databank will track whether participants have agreed to be re-contact or not to determine whether future data/sample collection is possible.**

In addition, re-contact is encouraged, where appropriate, to participants who were minor at the time of enrollment, for re-consent purposes (in accordance with local policy), to confirm that the participants choose to maintain their data in the HostSeq Databank.

12. Return of research results and incidental findings

The information resulting from research using HostSeq data is for scientific purposes only. This information will not be entered in the participant's medical file or communicated to the participant or to unauthorized third parties.

Incidental findings, which are findings that fall outside the scope of a particular research project, will not be returned to adult participants or to their physicians, as they will not be clinically relevant.

In the case of minor participants material actionable findings will always be returned for conditions that are actionable immediately or during childhood. These findings will be communicated by HostSeq to the contributing collection, who will be responsible for communicating such findings to the participant. The participant's parent/guardian will not be offered the choice of whether these types of actionable findings are returned or not, and must receive the findings, as the best interest of the minor participant prevails. Material actionable findings for conditions that are only actionable only in adulthood will not be returned to the participant's parent/guardian. Rather, at the time of recontact at majority the participant who was minor at the time of initial participant will be asked his or her preference on the communication of incidental findings.

13. Results and intellectual property

Intellectual Property rights may not be claimed on the data stored in the HostSeq Databank. Intellectual Property rights on derived data should not impede data usage by the researchers accessing the shared resource.

Patent protection will not be sought for any of the gene discoveries, functional assays, or scientific approaches developed as a result of the use of HostSeq data. HostSeq believes that open science has the most rapid impact for patients. This does not mean that eventual products, drugs or tests may not be commercialized by approved Third-Party researchers or Collection-affiliated Researchers.

Appendix A: Core Research Consent Elements for prospective collection of datasets

Consent to the collection of tissue samples by local sites, followed by the genetic analysis of tissue samples and sharing of research data through the HostSeq Databank, should be obtained by researchers at the local site. This consent can be obtained from the tissue donor (first person consent) or, where applicable, by obtaining consent from their legally authorized representative (substituted consent). Due to the nature of COVID-19, in many cases, consent may need to be obtained verbally, and necessary adjustment to the consent form and consent guidelines should be implemented, as recommended by the local site.

Prospective cohort consent forms should minimally include the following **core consent elements** to foster broad data sharing through the HostSeq Databank[6]:

	If a researcher at a local site wishes to deposit datasets in **HostSeq COVID-19 controlled-access Databank,** research consent should be obtained for:
Research data	Whole genome sequencing of the sample and the ongoing collection of clinical data from participant's medical records/chart, administrative databases, etc.
International sharing	International sharing of genetic and clinical data
Future research use	Future health research on COVID-19 and other health outcomes
Commercial use	Use of genetic and clinical data for commercial purposes
Controlled access	Sharing of genetic and clinical data through a controlled-access mechanism
Storage on cloud servers in Canada	Storage of genetic and clinical data in the HostSeq Databank, on centralized Canadian cloud servers
Duration of storage	Indefinite storage of genetic and clinical data
Data withdrawal	Not possible to withdraw data that has already been distributed and used
Re-identification	Low risk that the participant could be re-identified in the future
Option for recontact of participants (not mandatory, strongly recommended)[7]	Optional recontact of participants (yes/no)

[6] Examples of appropriate consent language for each of these core elements are provided in Appendix A.

[7] HostSeq strongly recommends that contributing collections include an option to re-contact participants in their consent forms. HostSeq will accept datasets from participants who have agreed or refused re-contact, however, the databank will track whether participants have agreed to be re-contact or not in order to determine whether future data/sample collection is possible.

If any of the items listed above are not included in the consent documents, datasets should not be deposited as-is in the HostSeq Databank without obtaining appropriate approvals from your local research ethics committee.

Appendix B: Example of consent clauses for the prospective Core Consent Elements

	Core Consent Elements	Example of consent clause language:
Research data	Whole genome sequencing of the sample and the ongoing collection of clinical data from participant's medical records/chart, administrative databases, etc.	You are being asked to consent for the whole genome sequencing of the DNA from your sample and to the storage of your **genetic data** in the HostSeq COVID-19 Databank, to be used for research purposes. You will also be asked to provide **clinical data** which includes some personal information about yourself and your health, as well information such as your age, ethnicity, and family's health history. We will also request health information about you from your family doctor and from other institutions or registries that may have your health information, for example, [*where applicable, include any relevant governmental/administrative health data repository in your province*]. We may get research data from other studies that you were involved in or future studies, if you agree.
International sharing	International sharing of genetic and clinical data	The HostSeq Databank will share your genetic and clinical data with approved researchers in Canada and around the world (which may include national and international researchers from academia, charitable organisations, hospitals, and 'for-profit' private companies, such as drug companies).
Future research use	Future health research on COVID-19 and related disorders	Your stored genetic and clinical data will be shared by the CGen Databank for future research on COVID-19 and other health outcomes.
Commercial use	Use of genetic and clinical data for commercial purposes	It is possible that future research conducted using your data combined with data from others will eventually lead to the development of new diagnostic tests, new drugs or other commercial products. If this happens, you will not receive any part of the profits from such products. The rights to the commercial products will belong to the sponsor, collaborators or future unknown researchers.
Controlled access	Sharing of genetic and clinical data through a controlled-access mechanism	Your data will only be shared with researchers approved by CGen after review by a Data Access Compliance Office (DACO). This DACO will verify, among other criteria, that the proposed research use is in conformity with the objectives of the HostSeq Databank, and that the research team applying for access has obtained the proper research ethics approval. Approved researchers will sign agreements. These agreements will control how the data will be used.
Storage on cloud servers in Canada	Storage of genetic and clinical data in the HostSeq Databank, on	Data in the HostSeq Databank is under the responsibility of CGen. CGen is a Canadian, federally funded national platform for genome sequencing and analysis.

	centralized Canadian servers	Data in the HostSeq Databank might be stored on cloud servers (i.e. online), based in Canada.
Duration of storage	Indefinite storage of genetic and clinical data	The data stored in the HostSeq Databank will be kept indefinitely, or until it is no longer useful for research.
Data withdrawal	Not possible to withdraw data that has already been distributed and used	If you decide to withdraw from the HostSeq Databank, your data will no longer be shared, and no new data will be collected. If you decide to withdraw from the study, data stored in the HostSeq Databank will be destroyed as of the time of your notification. However, it may be impossible to withdraw the results once they have been compiled with the results of others participating in the study or if they have been published. Furthermore, if some of the data have been shared with other researchers, it may not be possible to remove this part of the data. In these cases of total withdrawal being impossible, your identity will still be protected.
Re-identification	Low risk that the participant could be re-identified in the future	There is always a small risk that your data may lead to you being re-identified one day. Genetic information is unique to every person, just like a fingerprint. This means that theoretically you can be identified by your genetic code. However, this is not easy to do. As technology advances, there may be new ways of linking data back to you that we cannot foresee today, in spite of the strict security measures. The potential re-identification or unintentional release of your genetic and clinical research data could lead to loss of privacy and to possible future discrimination against you or your biological relatives.
Recontact	Optional recontact of participants (yes/no) (not mandatory, strongly recommended)	I agree to be re-contacted to update my personal information, to obtain additional health information, or, to be invited to participate in new research projects: _____**Yes** _____ **initials** _____**No** _____**initials**

Appendix C: Assessment of consent elements for inclusion of datasets from retrospective collections

The steps below aim to provide guidance on determining whether samples or datasets from retrospective collections are suitable for inclusion in the HostSeq COVID-19 Databank. Data submitted to the HostSeq COVID-19 Databank will be maintained in a controlled access database. Retrospective datasets include for example:

- Datasets that could be generated from already collected tissue samples (e.g. archival samples, samples collected in the context of another research project/biobank, samples collected, etc.);
- Datasets that were generated before the creation of the HostSeq COVID-19 Databank.

Retrospective consent materials may have used different language and in some cases may be ambiguous, or be silent, as to sample and data sharing and their potential uses. To help researchers from local sites determine whether samples and/or data can be used for the HostSeq COVID-19 Database, the following assessment tool was developed, based on the principles found in the Global Alliance for Genomics and Health's Consent Policy[8] and other best practice guidelines[9].

***Question 1*:** Does the consent form permit genetic sequencing of the tissue sample collected?

- If the question is **Not Applicable** because the tissue sample has already been analyzed and genetic data has already been derived with appropriate authorization, please proceed to ***Question 2***.
- If the answer is **Yes**, please proceed to ***Question 2***.
- If the answer is **No**, please proceed to ***Question 3* for re-consent or consent waivers**.

Question 2:

Does the consent form pertaining to datasets from retrospective collections indicate that:	***Yes***	***No***
a) Genetic analyses will be undertaken on the tissue sample and relevant clinical data will be collected?		
b) Data will be shared internationally?		
c) Data will be used for future health research on COVID-19 and related disorders, or for broader health/biomedical research uses?		
d) Data may be transferred and stored outside of the institution where it was collected and be made available to researchers through that external platform?		
e) Data will be stored for an indefinite period of time?		

[8] Global Alliance for Genomics and Health: Consent Policy (September 2019), available online at: https://www.ga4gh.org/wp-content/uploads/GA4GH-Final-Revised-Consent-Policy_16Sept2019.pdf

[9] For example, Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans – TCPS2 (2018), Section 5.5A (Consent and secondary use of identifiable information for research purposes) and 12.3A (Consent and secondary use of human biological materials for research purposes).

f) It may not be possible to withdraw data that has already been distributed and used by researchers?	
g) There is a minimal risk of re-identification in the future?	
h) Data can be used for commercial purposes?	
i) Genetic and clinical data will be shared through a controlled-access mechanism, meaning that research applicants are required to apply to an access committee to request access to the data, or similar process/consent language?	

- If **you have answered <u>yes to all of the above</u>**, core consent elements are deemed to be met to submit datasets to the HostSeq COVID-19 Databank. (In case of doubt or ambiguity in the consent language, please consult your local research ethics committee for guidance).
- If **you have answered <u>no to any of the above</u>**, please proceed to ***Question 3*** to determine if re-consent of participants is possible, or whether a consent waiver should be sought from your institutional ethics committee to deposit data in the HostSeq COVID-19 Databank.

Question 3:

Re-contact / re-consent	***Yes***	***No***
a) Does your consent form allow for re-contact of participants?		
b) Is it feasible for you to re-contact and re-consent your participants for inclusion of their data in the HostSeq COVID-19 Databank?		

- If **you have answered <u>yes to both questions</u>**, please re-contact and re-consent the participants and include the HostSeq COVID-19 Databank Consent Elements for prospective collections in your consent material.
- If **you have answered <u>no to either question</u>**, please proceed to ***Question 4***.

Question 4:

Requesting a waiver of consent	***Yes***	***No***
a) Is it possible for you to apply to your local ethics committee (or equivalent) to obtain a waiver of consent requirement in order to deposit your dataset in the HostSeq COVID-19 Databank? (See TCPS, article 5.5A and 12.3A)		

- If **you have answered <u>yes</u>**, please request and obtain a consent waiver you're your institutional research ethics committee, according to your local procedures.
- If **you have answered <u>no</u>**, your data cannot by deposited in the HostSeq COVID-19 Databank.

****

Appendix D: Data Access Compliance Office (DACO) Terms of Reference

1. Role
The role of the HostSeq Independent Data Access Committee (DACO) is to receive access requests for projects proposing to use controlled-access data. The DACO assesses the feasibility of the proposal, reviews and approves, conditionally approves or denies the request, according to criteria set out in Appendix E.

2. Term
These Terms of Reference are effective from [TBD].

3. Membership
The Committee shall be composed of a minimum of five (5) members, independent from CGen (the sponsor), including:

- At least one (1) bioinformatics researcher [voting member];
- At least one (1) expert in the technical infrastructure of the HostSeq Databank [voting member];
- At least one (1) expert on the legal/ethical aspect of genomic research, data sharing, privacy and data protection [voting member];
- One (1) investigator representing the HostSeq Databank [non-voting member];
- At least one (1) patient/participant representative or member of the public [voting member].

For certain access applications, the DACO Chair may call upon outside experts. Such expert will not be a voting member on the Committee, but invited to provide background expertise required to review the application.

4. Responsibilities
The membership of the DACO commits to:

- Attend all scheduled DACO meetings or, in case of absence, providing detailed notes and decisions to the DACO ahead of the meeting;
- Share all communications and information across all DACO members;
- Make timely decisions regarding data access requests.

Members of the DACO expect:

- That each DACO member is provided with complete, accurate and meaningful information in a timely manner prior to each meeting;
- To be given reasonable time to make key decisions.

5. Meetings
The Committee will appoint a Chair to preside over meetings and to review applications submitted for expedited review.

A meeting quorum is of 3 members of the DACO, in person or by video/phone.

Decisions are made by consensus (*i.e.* members are satisfied with the decision even though it may not be their first choice). If consensus is not possible, the DACO Chair makes the final decision.

Meeting agendas and minutes are provided by an Access Officer this includes:

- Preparing agendas and supporting papers (including list of access applications to be reviewed);
- Preparing meeting notes and information;
- Recording decisions to approve, conditionally approve or deny access requests.

Meetings are held monthly by tele- or video- conference.

6. Amendment, Modification or Variation

These Terms of Reference may be amended, varied or modified in writing after consultation and agreement by the DACO.

Appendix E: Steps to request access to controlled-access datasets

Controlled-access datasets- Regular review

In the case of a request by **a Third-Party Researcher to access controlled-access data**, a project-specific application needs to be submitted to the HostSeq DACO, for review.

The following steps summarize the controlled-access application process:

1. The Principle Investigator (PI) must submit a completed Application Form requesting access to controlled-access data to the DACO. A new application form must be submitted for each new study. As part of the application, the PI should list all other research team members who require access to the data (ex: researchers, employees, laboratory personnel, etc.).
2. As part of its review, the DACO may consider the following information:
   - Compatibility of the proposal with the objectives of the HostSeq Databank;
   - Description of the dataset requested and availability within the HostSeq Databank;
   - Justification that the project requires access to HostSeq datasets;
   - Qualifications of the PI to undertake the proposed project;
   - Number and names of research team members that will have access to data and rationale for their access;
   - Scientific merit and feasibility of the proposed project;
   - Proof of local ethics committee approval for the proposed project (or justification as to why such approval is not required); and
3. Upon approval by the DACO, a Data Access Agreement is entered into between the PI and the SickKids.
4. Controlled access to HostSeq Databank is granted for a period of 1 year (which may be renewed, upon request).
5. Approved PIs may contact the DACO for any amendment to the initial application;

Controlled-access datasets- Expedited review

In the case of a request by **Collection-Affiliated Researcher requesting access to controlled access data from collections other than their own**, a project-specific application needs to be submitted to the HostSeq DACO, for expedited review by the DACO Chair.

The following steps summarize the controlled-access expedited review process:

1. The Principle Investigator (PI) must submit a completed Application Form requesting access to controlled-access data to the DACO Chair. A new application form must be submitted for each new study. As part of the application, the PI should list all other research team members who require access to the data (ex: researchers, employees, laboratory personnel, etc.).
2. As part of their review, a DACO Chair will review the following information:
   - Compatibility of the proposal with the objectives of the HostSeq Databank;
   - Description of the dataset requested and availability within the HostSeq Databank;
   - Scientific merit and feasibility of the proposed project;
   - Proof of local ethics committee approval for the proposed project (or justification as to why such approval is not required);
3. Upon approval by the DACO Chair, or delegate, a Data Access Agreement is entered into between the PI and SickKids.
4. Controlled access to HostSeq Databank is granted for a period of 1 year (which may be renewed, upon request).
5. Approved PIs may contact the DACO Chair for any amendment to the initial application.

No	Version	Date	Description
1	v.1.0	04 August 2020	-First draft approved by the HostSeq Implementation Committee